

The relation between frequentist confidence intervals and Bayesian credible intervals

S.I.Bitioukov and N.V.Krasnikov

INR RAS, Moscow 117312

November 15, 2012

Abstract

We investigate the relation between frequentist and Bayesian approaches. Namely, we find the “frequentist” Bayes prior $\pi_f(\lambda, x_{obs}) = -\frac{\int_{-\infty}^{x_{obs}} \frac{\partial f(x, \lambda)}{\partial \lambda} dx}{f(x_{obs}, \lambda)}$ (here $f(x, \lambda)$ is the probability density) for which the results of frequentist and Bayes approaches to the determination of confidence intervals coincide. In many cases (but not always) the “frequentist” prior which reproduces frequentist results coincides with the Jeffreys prior.

One of the standard problems in statistics [1] is an estimation of the values of unknown parameters in the probability density. There are two methods to solve this problem - the frequentist and the Bayesian.

In this paper we investigate the relation between frequentist and Bayesian approaches. Namely, we find the “frequentist” Bayes prior $\pi_f(\lambda, x_{obs}) = -\frac{\int_{-\infty}^{x_{obs}} \frac{\partial f(x, \lambda)}{\partial \lambda} dx}{f(x_{obs}, \lambda)}$ (here $f(x, \lambda)$ is the probability density) for which the results of frequentist and Bayes approaches to the determination of confidence intervals coincide. In many cases (but not always) the “frequentist” prior coincides with the Jeffreys prior. Note that in ref.[2] the relation between frequentist confidence intervals and Bayesian credible intervals has been found for probabilities densities of the special type $f(x, \lambda) = \Phi(x - \lambda)$ and $f(x, \lambda) = \frac{1}{\lambda} F(\frac{x}{\lambda})$.

As an example consider the case of random continuous observable $-\infty < x < \infty$ with the probability density $f(x, \lambda)$ ¹.

In Bayesian method [1, 3] due to Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

the probability density for unknown parameter λ is determined as

$$p(\lambda|x_{obs}) = \frac{f(x_{obs}, \lambda)\pi(\lambda)}{\int_{-\infty}^{+\infty} f(x_{obs}, \lambda')\pi(\lambda')d\lambda'}. \quad (2)$$

Here x_{obs} is the observed value of the random variable x and $\pi(\lambda)$ is the prior function. In general the prior function $\pi(\lambda)$ is not known that is the main problem of the Bayesian approach. Formula (2) reduces the statistics problem to the probability problem. The probability that parameter λ lies in the interval $\lambda_{down} \leq \lambda \leq \lambda_{up}$ is ²

$$P(\lambda_{down} \leq \lambda \leq \lambda_{up}) = \int_{\lambda_{down}}^{\lambda_{up}} p(\lambda|x_{obs})d\lambda = 1 - \alpha(\lambda_{up}, \lambda_{down}|x_{obs}), \quad (3)$$

where

$$\alpha(\lambda_{up}, \lambda_{down}|x_{obs}) = \beta(\lambda_{up}|x_{obs}) + \gamma(\lambda_{down}|x_{obs}), \quad (4)$$

$$\beta(\lambda_{up}|x_{obs}) = \int_{\lambda_{up}}^{\infty} p(\lambda|x_{obs})dx, \quad (5)$$

¹Here $\infty < \lambda < -\infty$ is some unknown parameter and $\int_{-\infty}^{\infty} f(x, \lambda)dx = 1$.

²Usually $\alpha(\lambda_{up}, \lambda_{down}|x_{obs})$ is taken nondependent on $\lambda_{up}, \lambda_{down}$ and equal to 0.05.

$$\gamma(\lambda_{down}|x_{obs}) = \int_{-\infty}^{\lambda_{down}} p(\lambda|x_{obs})d\lambda = 1 - \beta(\lambda_{down}|x_{obs}) , \quad (6)$$

The solution of the equation (3) is not unique. The most popular are the following options [1]:

1. $\lambda_{down} = -\infty$ - upper limit.
2. $\lambda_{up} = \infty$ - lower limit.
3. $\int_{-\infty}^{\lambda_{down}} p(\lambda|x_{obs})d\lambda = \int_{\lambda_{up}}^{\infty} p(\lambda|x_{obs})d\lambda = \frac{\alpha}{2}$ - symmetric interval.
4. The shortest interval - $p(\lambda|x_{obs})$ inside the interval is bigger or equal to $p(\lambda|x_{obs})$ outside the interval.

outside the interval.

In frequentist approach the Neyman belt construction [4](see Fig. 1) is used for the determination of the confidence intervals.

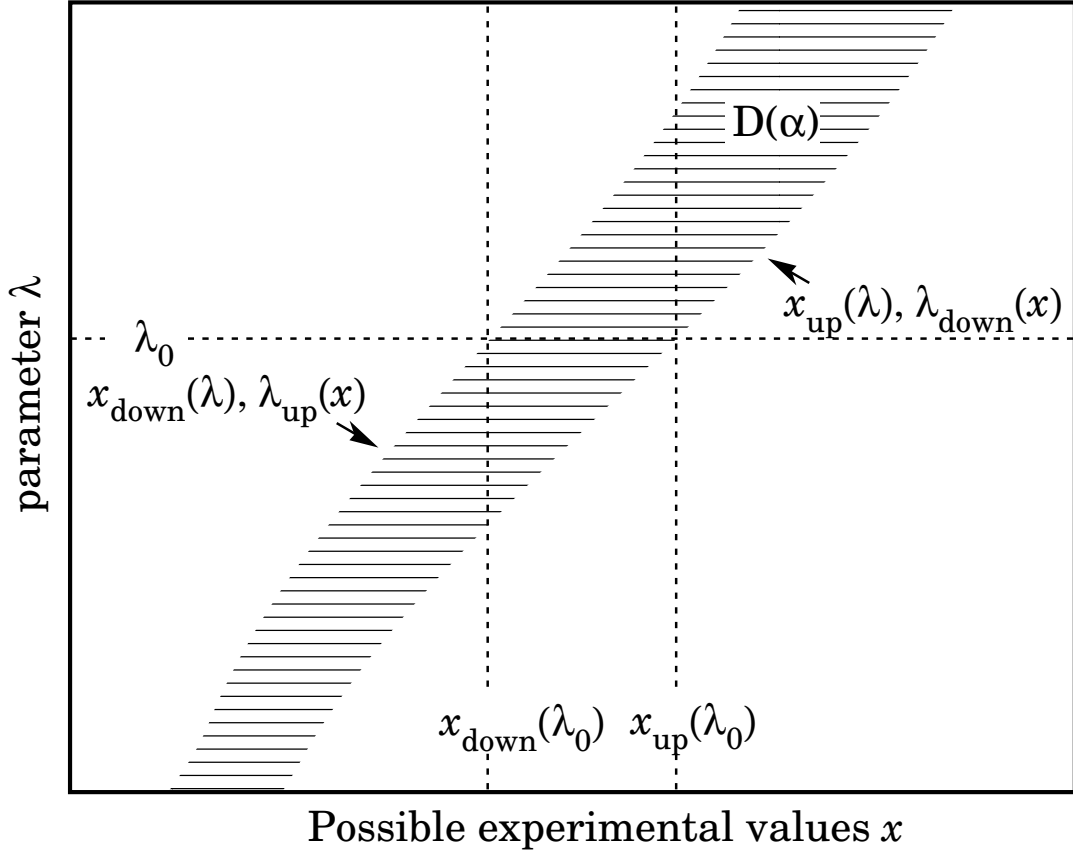


Figure 1: Neyman belt construction

Namely, we require that

$$\int_{x_{down}(\lambda)}^{x_{up}(\lambda)} f(x, \lambda)dx = 1 - \alpha , \quad (7)$$

or

$$\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx + \int_{-\infty}^{x_{down}(\lambda)} f(x, \lambda) dx = \alpha. \quad (8)$$

It should be stressed that in general case both $\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx$ and $\int_{-\infty}^{x_{down}(\lambda)} f(x, \lambda) dx$ can depend on λ but their sum does not depend on λ (see eq.(8)).

The Neyman equations for the determination of upper and lower limits λ_{up} and λ_{down} on parameter λ have the form

$$\int_{x_{obs}}^{x_{up}(\lambda_{up})} f(x, \lambda_{up}) dx = 1 - \alpha, \quad (9)$$

$$\int_{x_{down}(\lambda_{down})}^{x_{obs}} f(x, \lambda_{down}) dx = 1 - \alpha \quad (10)$$

and they determine the confidence interval of possible values

$$\lambda_{down} \leq \lambda \leq \lambda_{up} \quad (11)$$

of the parameter λ at the $(1 - \alpha)$ confidence level.

In this paper we shall consider the case when $\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx$ does not depend on λ . For instance, the options $x_{up}(\lambda) = \infty$, $x_{down}(\lambda) = -\infty$ and $\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx = \int_{-\infty}^{x_{down}(\lambda)} f(x, \lambda) dx = \frac{\alpha}{2}$ correspond to the cases of upper limit on λ , lower limit on λ and symmetric interval correspondingly. As a consequence of the eqs.(8-11) and our assumption on nondependence of $\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx$ on λ we find that

$$\alpha = \int_{-\infty}^{x_{obs}} f(x, \lambda_{up}) dx + \int_{x_{obs}}^{\infty} f(x, \lambda_{down}) dx. \quad (12)$$

To find the relation between frequentist and Bayesian approaches we have to find the prior for which the formulae (4-6) and (12) coincide. Namely, we require that

$$\int_{-\infty}^{\lambda_{down}} p(\lambda | x_{obs}) d\lambda + \int_{\lambda_{up}}^{\infty} p(\lambda | x_{obs}) d\lambda = \int_{-\infty}^{x_{obs}} f(x, \lambda_{up}) dx + \int_{x_{obs}}^{\infty} f(x, \lambda_{down}) dx. \quad (13)$$

The solution of eq.(13) is

$$p(\lambda | x_{obs}) = - \int_{-\infty}^{x_{obs}} \frac{\partial f(x, \lambda)}{\partial \lambda} dx. \quad (14)$$

Formulae (13,14) demonstrate the equivalence of the frequentist approach and the Bayes approach with the prior function

$$\pi_f(\lambda | x_{obs}) = - \frac{\int_{-\infty}^{x_{obs}} \frac{\partial f(x, \lambda)}{\partial \lambda} dx}{f(x_{obs}, \lambda)}. \quad (15)$$

Note that in the limit $\lambda_{down} \rightarrow -\infty$ and $\lambda_{up} \rightarrow \infty$ full probability must be equal to one, namely

$$\lim_{\lambda_{down} \rightarrow -\infty, \lambda_{up} \rightarrow \infty} \int_{-\infty}^{x_{obs}} [f(x, \lambda_{down}) - f(x, \lambda_{up})] dx = 1. \quad (16)$$

As a consequence of the equation (16) we find that

$$\lim_{\lambda_{down} \rightarrow -\infty} \int_{-\infty}^{x_{obs}} f(x, \lambda_{down}) dx = 1, \quad (17)$$

$$\lim_{\lambda_{up} \rightarrow \infty} \int_{-\infty}^{x_{obs}} f(x, \lambda_{up}) dx = 0. \quad (18)$$

Consider several examples. For the probability density

$$f(x, \lambda) = \Phi(x - \lambda) \quad (19)$$

as a consequence of the formulae (14-15) we find that

$$p_f(\lambda, x_{obs}) = \Phi(x_{obs} - \lambda), \quad (20)$$

$$\pi_f(\lambda, x_{obs}) = 1. \quad (21)$$

Note that for the distribution (19) the Jeffreys prior [5] $\pi(\lambda) \sim \sqrt{[\int_{-\infty}^{\infty} f(x, \lambda) (\frac{\partial \ln(f(x, \lambda))}{\partial \lambda})^2 dx]} \sim \text{const}$ does not depend on λ , i.e. for the distribution (19) the frequentist approach is equivalent to the Bayes approach with the Jeffreys prior $\pi(\lambda) = \text{const}$ [2].

Consider the case with the parameter $\lambda \geq b$. Here b is some fixed number³. Such situation arises when we measure signal s in the presence of nonzero background b and $\lambda = b + s$, $b \geq 0$, $s \geq 0$. The parameter λ lies in the interval $b \leq \lambda < \infty$. The direct use of the formulae (12,13) leads to the inconsistency. Namely, we find that the probability

$$P(b < \lambda < \infty) = \int_b^{\infty} \Phi(x_{obs} - \lambda) d\lambda < 1 \quad (22)$$

that contradicts to the postulate that the full probability must be equal to 1. At the frequentist language the inequality (22) is the consequence of the fact that

$$\int_{x_{obs}}^{\infty} f(x, \lambda_{down} = b) dx \neq 0. \quad (23)$$

³In ref. [6] normal distribution $N(x, \mu, \sigma^2)$ with additional constraint $\mu \geq 0$ has been studied.

To obtain the correct solution we must use the language of the conditional probabilities. Really, the probability that parameter λ lies in the interval $\lambda_0 \leq \lambda \leq \lambda_0 + d\lambda$ is equal to $\Phi(x_{obs} - \lambda_0)d\lambda$. The probability that parameter λ lies in the interval $\lambda_0 \leq \lambda \leq \lambda_0 + d\lambda$ provided $\lambda \geq b$ is determined by the formula of the conditional probability

$$P(\lambda_0 \leq \lambda \leq \lambda_0 + d\lambda | \lambda \geq b) = \frac{P(\lambda_0 \leq \lambda \leq \lambda_0 + d\lambda)}{P(\lambda \geq b)} = \frac{\Phi(x_{obs} - \lambda_0)d\lambda}{\int_b^\infty \Phi(x_{obs} - \lambda)d\lambda}. \quad (24)$$

So we see that condition $\lambda \geq b$ leads to the appearance of additional factor $\int_b^\infty \Phi(x_{obs} - \lambda)d\lambda$ in the denominator of the formula (24). This factor restores the requirement that full probability $P(b_0 \leq \lambda < \infty) = 1$. For instance, the probability that signal s is less than s_0 is determined by the formula

$$P(s \leq s_0) = \frac{\int_0^{s_0} \Phi(x_{obs} - b - s)ds}{\int_b^\infty \Phi(x_{obs} - \lambda)d\lambda} \quad (25)$$

and it coincides with the corresponding formula of the CL_s method [7, 8].

For the probability density

$$F(x, \lambda) = \frac{1}{\lambda} \Phi\left(\frac{x}{\lambda}\right) \quad (26)$$

we find that

$$p_f(\lambda, x_{obs}) = \frac{x_{obs}}{\lambda^2} \Phi\left(\frac{x_{obs}}{\lambda}\right), \quad (27)$$

$$\pi_f(\lambda, x_{obs}) = \frac{x_{obs}}{\lambda}. \quad (28)$$

Again in this case the prior (28) coincides with the Jeffreys prior [2].

Consider the probability density⁴

$$f(x, \lambda) = \theta(x)P([x], \lambda), \quad (29)$$

where

$$P([x], \lambda) = \frac{\lambda^{[x]}}{[x]!} e^{-\lambda} \quad (30)$$

is the Poisson distribution and $[x]$ is an integer part of x (for instance, $[2.33] = 2$). Using formulae (14,15) one can find that for the probability density (29) and $x_{obs} \geq 1$ ⁵

$$p_f(x_{obs}, \lambda) = P([x_{obs}] - 1, \lambda) + (x_{obs} - [x_{obs}])(-P([x_{obs}] - 1, \lambda) + P([x_{obs}], \lambda)), \quad (31)$$

⁴For the probability density (29) $\int_{-\infty}^{\infty} f(x, \lambda)dx = \sum_{n=0}^{\infty} P(n, \lambda) = 1$.

⁵For $x_{obs} < 1$ $p_f(x_{obs}, \lambda) = P(0, \lambda)$ and $\pi_f(\lambda, x_{obs}) = 1$.

$$\pi_f(\lambda, x_{obs}) = \frac{[x_{obs}]}{\lambda} + (x_{obs} - [x_{obs}])(-\frac{[x_{obs}]}{\lambda} + 1). \quad (32)$$

The Jeffreys prior for the distribution function (29) coincides with the Jeffreys prior for Poisson distribution and it is proportional to $\pi(\lambda) \sim \frac{1}{\sqrt{\lambda}}$. So we see that for the probability density (29) the “frequentist prior” (32) and the Jeffreys prior are different.

Consider now the case when $\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx$ depends on λ . Instead of $x_{up}(\lambda)$ and $x_{down}(\lambda)$ we can find another $x_{up}^{\prime}(\lambda)$ and $x_{down}^{\prime}(\lambda)$ such that $x_{up}(\lambda_{up}) = x_{up}^{\prime}(\lambda_{up}) = x_{obs}$, $x_{down}(\lambda_{down}) = x_{down}^{\prime}(\lambda_{down}) = x_{obs}$ and

$$\int_{-\infty}^{x_{up}^{\prime}(\lambda)} f(x, \lambda) dx = \int_{-\infty}^{x_{obs}} f(x, \lambda_{up}) dx, \quad (33)$$

$$\int_{x_{down}^{\prime}(\lambda)}^{\infty} f(x, \lambda) dx = \int_{x_{obs}}^{\infty} f(x, \lambda_{down}) dx. \quad (34)$$

For the confidence intervals defined by new trajectories $x_{up}^{\prime}(\lambda)$ and $x_{down}^{\prime}(\lambda)$ the Neyman equations have the form

$$\int_{x_{down}^{\prime}(\lambda)}^{x_{up}^{\prime}(\lambda)} f(x, \lambda) dx = 1 - \alpha', \quad (35)$$

$$\alpha' = \int_{-\infty}^{x_{obs}} f(x, \lambda_{up}) dx + \int_{x_{obs}}^{\infty} f(x, \lambda_{down}) dx. \quad (36)$$

and the equations (14,15) are valid. So we see that for the Neyman belt construction with $x_{up}^{\prime}(\lambda)$, $x_{down}^{\prime}(\lambda)$ and the integral $\int_{x_{down}^{\prime}(\lambda)}^{\infty} f(x, \lambda) dx$ nondependent on λ the frequentist approach with α' given by the expression (36) and the Bayesian approach with prior function (14) coincide.

In conclusion let us formulate our main result. For the particular case when $\int_{x_{up}(\lambda)}^{\infty} f(x, \lambda) dx$ does not depend on λ we have found the “frequentist” Bayes prior $\pi_f(\lambda, x_{obs}) = -\frac{\int_{-\infty}^{x_{obs}} \frac{\partial f(x, \lambda)}{\partial \lambda} dx}{f(x_{obs}, \lambda)}$ for which the results of the frequentist and the Bayes approaches to the determination of confidence intervals coincide. In many cases (but not always) the prior which reproduces frequentist results coincides with the Jeffreys prior.

This work has been supported by RFBR grant N 10-02-00468.

References

- [1] As a review, see for example:
F.James, “Statistical methods in experimental physics”, 2nd ed.,,
World Scientific, 2006.
- [2] E.T.Jaynes, “Confidence intrvals vs Bayesian intervals”, in Foundations of
Probability Theory, Statistical Inference and Statistical Theories of Science
(W.L.Harper and C.A.Hooker, eds), pp.175-257, 1976.
- [3] See, for example:
G.D’Agostini, “Bayesian Reasoning in Data Analysis, a Critical Introduction”,
World Scientific, Hackensack, NJ, 2003.
- [4] Y.Neyman, Philos. Trans. R. Soc.London Sect. **A236** 333 (1937).
- [5] Harold Jeffreys, “Theory of Probability”, 3 rd ed.,
(Oxford University Press, Oxford, 1961).
- [6] G.J.Feldman and R.D.Cousins, Phys.Rev.**D57** 3873 (1998).
- [7] T.Junk, Nucl.Instrum.Meth. **A434** 435 (1999).
- [8] A.L.Read, “Modified frequentist analysis of search results(the CL_s method)”,
CERN Yellow Report **CERN-2000-005** 81 (2000).